AFRL-OSR-VA-TR-2013-0563

ADAPTIVE KERNEL BASED MACHINE LEARNING METHODS

YUESHENG XU

SYRACUSE UNIVERSITY

10/15/2013
Final Report

DISTRIBUTION A: Distribution approved for public release.

AIR FORCE RESEARCH LABORATORY
AF OFFICE OF SCIENTIFIC RESEARCH (AFOSR)/RSL
ARLINGTON, VIRGINIA 22203
AIR FORCE MATERIEL COMMAND

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 10/10/2012	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* 07/1/09 - 06/30/12

4. TITLE AND SUBTITLE
Adaptive Kernel Based Machine Learning Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-09-1-0511

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Yuesheng Xu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Syracuse University
Office of Sponsored Programs
Syracuse, NY 13244-1200

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/RTC
875 North Randolph Street, 4036
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Publically Releasable

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Research results obtained from this project address the kernel selection problem in machine learning. Specifically, motivated from the need of updating the current operator-valued reproducing kernel in multi-task learning when underfitting or overfitting occurs, we studied the construction of a refinement kernel for a given operator-valued reproducing kernel such that the vector-valued reproducing kernel Hilbert space of the refinement kernel contains that of the given kernel as a subspace. We also developed a complete characterization of multi-task finite rank kernels in terms of the positivity of what we call its associated characteristic operator.

15. SUBJECT TERMS
kernel based machine learning method, multi-task learning

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Yuesheng Xu
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)* 315-443-1575

Final Report of "Adaptive Kernel Based Machine Learning Methods"

PI: Yuesheng Xu, Syracuse University

Sponsor Award No.: FA9550-09-1-0511

Abstract

During the support period July 1, 2011 - June 30, 2012, seven research papers were published. They consist of three types:

- Research that directly addresses the kernel selection problem in machine learning [1, 2].
- Research that closely relates to the fundamental issues of the proposed research of this grant [3, 4, 5, 6].
- Research that is in the general context of computational mathematics [7].

Paper [1] studies the construction of a refinement kernel for a given operator-valued reproducing kernel such that the vector-valued reproducing kernel Hilbert space of the refinement kernel contains that of the given kernel as a subspace. The study is motivated from the need of updating the current operator-valued reproducing kernel in multi-task learning when underfitting or overfitting occurs. Numerical simulations confirm that the established refinement kernel method is able to meet this need. Various characterizations are provided based on feature maps and vector-valued integral representations of operator-valued reproducing kernels. Concrete examples of refining translation invariant and finite Hilbert-Schmidt operator-valued reproducing kernels are provided. Other examples include refinement of Hessian of scalar-valued translation-invariant kernels and transformation kernels. Existence and properties of operator-valued reproducing kernels preserved during the refinement process are also investigated.

Motivated by the importance of kernel-based methods for multi-task learning, we provide in [2] a complete characterization of multi-task finite rank kernels in terms of the positivity of what we call its associated characteristic operator. Consequently, we are led to establishing that every continuous multitask kernel, defined on a cube in an Euclidean space, not only can be uniformly approximated by multi-task polynomial kernels, but also can be extended as a multi-task kernel to all of the Euclidean space. Finally, we discuss the interpolation of multi-task kernels by multi-task finite rank kernels.

Multiscale collocation methods are developed in [3] for solving a system of integral equations which is a reformulation of the Tikhonov-regularized second-kind equation of an ill-posed integral equation of the first kind. This problem is closely related to regularization problems in machine learning. Direct numerical solutions of the Tikhonov regularization equation require one to generate a matrix representation of the composition of the conjugate operator with its original integral operator. Generating such a matrix is computationally costly. To overcome this challenging computational issue, rather than directly solving the Tikhonov-regularized equation, we propose to solve an equivalent coupled system of integral equations. We apply a multiscale collocation method with a matrix compression strategy to discretize the system of integral equations and then use the multilevel augmentation method to solve the resulting discrete system. A priori and a posteriori

parameter choice strategies are developed for thesemethods. The convergence order of the resulting regularized solutions is estimated. Numerical experiments are presented to demonstrate the approximation accuracy and computational efficiency of the proposed methods.

Although the computational method developed in the three papers [4, 5, 6] are for image processing, the ideas used in this development are suitable for solving regularization problems in machine learning.

The paper [4] introduces a proximity operator framework for studying the L1/TV image denoising model which minimizes the sum of a data fidelity term measured in the L1-norm and the total-variation regularization term. Both terms in the model are non-differentiable. This causes algorithmic difficulties for its numerical treatment. To overcome the difficulties, we formulate the total-variation as a composition of a convex function (the L1-norm or the L2- norm) and the first order difference operator, and then express the solution of the model in terms of the proximity operator of the composition. By developing a chain rule for the proximity operator of the composition, we identify the solution as fixed point of a nonlinear mapping expressed in terms of the proximity operator of the L1-norm or the L2-norm, each of which is explicitly given. This formulation naturally leads to fixed-point algorithms for the numerical treatment of the model. We propose an alternative model by replacing the non-differentiable convex function in the formulation of the total variation with its differentiable Moreau envelope and develop corresponding fixed-point algorithms for solving the new model. When partial information of the underlying image is available, we modify the model by adding an indicator function to the minimization functional and derive its corresponding fixed-point algorithms. Numerical experiments are conducted to test the approximation accuracy and computational efficiency of the proposed algorithms. Also, we provide a comparison of our approach to two state-of-the-art algorithms available in the literature. Numerical results confirm that our algorithms perform favorably, in terms of PSNR-values and CPU-time, in comparison to the two algorithms.

The goal of the paper [5] is to improve the computational performance of the proximity algorithms for the L1/TV denoising model. This leads us to a new characterization of all solutions to the L1/TV model via fixed-point equations expressed in terms of the proximity operators. Based upon this observation we develop an algorithm for solving the model and establish its convergence. Furthermore, we demonstrate that the proposed algorithm can be accelerated through the use of the componentwise GaussSeidel iteration so that the CPU time consumed is significantly reduced. Numerical experiments using the proposed algorithm for impulsive noise removal are included, with a comparison to three recently developed algorithms. The numerical results show that while the proposed algorithm enjoys a high quality of the restored images, as the other three known algorithms do, it performs significantly better in terms of computational efficiency measured in the CPU time consumed.

We propose in [6] a preconditioned alternating projection algorithm (PAPA) for solving the maximum a posteriori (MAP) emission computed tomography (ECT) reconstruction problem. Specifically, we formulate the reconstruction problem as a constrained convex optimization problem with the total variation (TV) regularization. We then characterize the solution of the constrained convex optimization problem and show that it satisfies a system of fixedpoint equations defined in terms of two proximity operators raised from the convex functions that define the TV-norm and the constrain involved in the problem. The characterization (of the solution) via the proximity operators that define two projection operators naturally leads to an alternating projection algorithm for finding the solution. For efficient numerical computation, we introduce to the alternating projection algorithm a preconditioning matrix (the EM-preconditioner) for the dense system matrix involved in the optimization problem. We prove theoretically convergence of the PAPA. In numerical experiments, performance of our algorithms, with an appropriately selected preconditioning matrix, is

compared with performance of the conventional MAP expectation-maximization (MAP-EM) algorithm with TV regularizer (EM-TV) and that of the recently developed nested EM-TV algorithm for ECT reconstruction. Based on the numerical experiments performed in this work, we observe that the alternating projection algorithm with the EM-preconditioner outperforms significantly the EM-TV in all aspects including the convergence speed, the noise in the reconstructed images and the image quality. It also outperforms the nested EM-TV in the convergence speed while providing comparable image quality.

We consider [7] the precondition of linear systems which resulted from the finite volume method (FVM) for elliptic boundary value problems. With the help of the interpolation operator from the trial space to the test space of the FVM and the operator induced by the FVM bilinear form, we show that both wavelet preconditioners and multilevel preconditioners designed originally for the finite element method (FEM) of a boundary value problem can be used to precondition the FVM of the same boundary value problem. We prove that such preconditioners ensure that the resulting coefficient matrix of the FVM has a uniformly bounded condition number. We present seven numerical examples to confirm our theoretical findings.

References

[1] H. Zhang, Y. Xu and Q. Zhang. Refinement of Operator-valued Reproducing Kernels. *Journal of Machine Learning Research*, 13 (2012) 91-136.

[2] J. Liu, C. A. Micchelli, R. Wang and Y. Xu Finite Rank Kernels for Multi-task Learning *Advances in Computational Mathematics*, DOI 10.1007/s10444-011-9244-x.

[3] Z. Chen, S. Ding, Y. Xu and H. Yang. Multiscale Collocation Methods for Ill-Posed Integral Equations via a Coupled System *Inverse Problems*, 28 (2012) 025006 (21pp).

[4] C. A. Micchelli, L. Shen, Y. Xu and X. Zeng. Proximity Algorithms for the L1/TV Image Denoising Model. *Advances in Computational Mathematics*, DOI 10.1007/s10444-011-9243-y.

[5] Q. Li, C. Micchelli, L. Shen and Y. Xu. A Proximity Algorithm Accelerated by GaussSeidel Iterations for L1/TV Denoising Models. *Inverse Problems*, 28 (2012) 095003 (20pp).

[6] A. Krol1, S.Li, L. Shen and Y. Xu Preconditioned Alternating Projection Algorithms for Maximum A Posteriori ECT Reconstruction. *Inverse Problems*, 28 (2012) 000000 (35pp), to appear.

[7] Y. Li, S. Shu, Y. Xu and Q. Zou. Multilevel Preconditioning for the Finite Volume Method. *Mathematics of Computation*, 81 (2012), 1399-1428.